Mint Centrality:
A Centrality Measure for the Bitcoin Transaction Graph

Beltran Borja Fiz Pontiveros
Sedan Group, SnT
University of Luxembourg
beltran.fiz@uni.lu

Mathis Steichen
Sedan Group, SnT
University of Luxembourg
mathis.steichen@uni.lu

Radu State
Sedan Group, SnT
University of Luxembourg
radu.state@uni.lu

Abstract—**In this work, we consider the graph of confirmed transactions in Bitcoin. Understanding this graph is essential to discern the different economic activities conducted by the pseudonymous actors. In addition to traditional graph analysis methods, new metrics need to be engineered specifically for the bitcoin transaction graph. Hence, we propose a new centrality measure named mint centrality. The measure uses the inherent tree structure of transactions in bitcoin and their relation to the corresponding set of coinbase transactions, and can be evaluated with linear complexity. We evaluate the mint centrality on the first 200,000 blocks of the public bitcoin blockchain.**

Index Terms—**Bitcoin, Centrality Measure, Transaction Graph**

1. Introduction

Bitcoin is a cryptocurrency and a decentralised payment system originally conceived by Satoshi Nakamoto in 2008 [1]. The bitcoin blockchain acts as a distributed ledger extended and validated by a peer-to-peer network of distrusting nodes that keeps track of the order of transactions.

In recent years the use of bitcoin as a payment platform for illicit goods has led to a rapid increase in sales [2]. Consequently, more research has been devoted to the analysis of the bitcoin transaction graph to track down these users, for example using network taint analysis [3]. This relies on the fact that bitcoin transactions are composed of inputs (previously unspent outputs) and new unspent outputs.

In order to prevent this type of analysis, bitcoin users stopped reusing addresses for multiple transactions, making it more difficult to track the total amount owned by a user. Moreover, services like bitcoin mixing or tumbler [4] allow users to merge multiple unspent outputs from multiple addresses into one transaction with multiple destinations. This allows the obfuscation of the trail from the sender to the receiver.

It is out of this idea of mixing bitcoin transactions from different addresses that the idea for the centrality measure proposed in this paper was born. By keeping track of the set of coinbase transactions that an unspent transaction originates from, we can keep track of the ”mixture level” of an unspent output. We can then use this value as a centrality measure in order to determine the importance of an address. The main contribution of this paper is the proposal of a novel centrality measure based on the innate structure of the bitcoin transaction graph.

2. Related Work

Bitcoin's underlying peer-to-peer is evaluated in [5] to determine the most influential nodes with respect to transaction broadcasting. Similarly, we attempt to discover influential addresses within the bitcoin network.

The authors in [6] performed a thorough analysis of the unspent transaction output set (UTXOs). They show that the set is growing rapidly and provide a taxonomy of different transaction types. UTXOs are of paramount importance to this work, as they relate accounts to transaction outputs and allow us to compute the mint centrality.

Regarding centrality measures, a lot of work across multiple fields has already dealt with the issue of comparing centrality metrics in numerous sample graphs of different properties [7], [8]. The authors in [9] focus on a thorough comparison of common centrality measures applied to networks. They show that by establishing a strong correlation between two centralities one could use a lower complexity centrality measure to approximate another slower centrality.

To the best of our knowledge no previous centrality measure has been designed specifically for the bitcoin transaction graph.

3. Background

The Bitcoin blockchain is extended with blocks through mining. The blocks contain new transactions that modify the state of the blockchain. Transactions are made of inputs and outputs. Inputs are outputs of one or more previous transactions that have not yet been spent, i.e. are still available. Outputs can be spent by the receiving account in a subsequent transaction. One particular kind of transaction are the coinbase transactions. They contain no inputs and reward the respective miners. Moreover, they provide the financial incentive for nodes to mine and are part of the

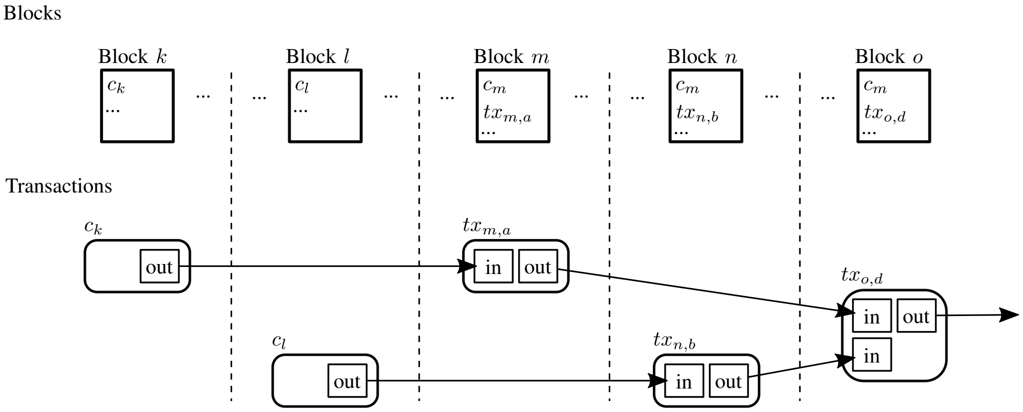

Figure 1. A representation of blocks, transactions and coinbases on the Bitcoin blockchain. Coinbase transactions are designated with c_i, where i is the block number, regular transactions are denoted with $tx_{i,j}$, where i is the block number and j is the number of the transaction within the block i. The inputs to a transaction are shown on its left side with the keyword 'in', while the outputs are depicted on the right with 'out'. The arrows show which output is used as an input to another transaction. For example, the output of c_k is used as an input to $tx_{m,a}$, while $tx_{o,d}$ uses outputs created by $tx_{m,a}$ and $tx_{n,b}$.

coin minting process in Bitcoin. New coins are exclusively created by these transactions.

Since inputs are previously unspent outputs, a DAG can be formed to represent transactions and their dependency on other transactions, those that created a given output. Furthermore, since only coinbase transactions create new coins, and by following the graph from output to input and so forth, every output can eventually be linked to several coinbase transactions from which the coins that are being spent originated. Extending on this, addresses can also be linked to coinbase transactions through the unspent transaction outputs (UTXOs) associated with them. More precisely, at any given point in time, any unspent transaction output belongs to an address and is linked to one or more coinbase transactions. Hence, there is also a relation between the address and the coinbase transactions that all of its UTXOs are associated with. This principle is the basis of the herein proposed mint centrality.

Figure 1 illustrates how transaction inputs and outputs can be linked to form a graph connecting outputs to coinbase transactions.

4. Methodology: mint centrality

In this section we define the proposed centrality and describe the implementation used for our experiments. In addition we discuss its time and space complexity.

Let's define the bitcoin transaction graph as a directed acyclic graph $G = (V, E)$. The set of vertices V denotes the transactions that occur on the Bitcoin blockchain. The edges E are the connections between transactions resulting from the discussion in Section 3. The direction of the edges is chosen as representing a dependency, hence pointing from a transaction using an output to a transaction creating it. It is in this manner that the graph is constructed. Consequently, only the coinbase transactions do not have a dependency, i.e. they do not have an outgoing edge pointing to another transaction. Figure 2 illustrates this graph on a small example.

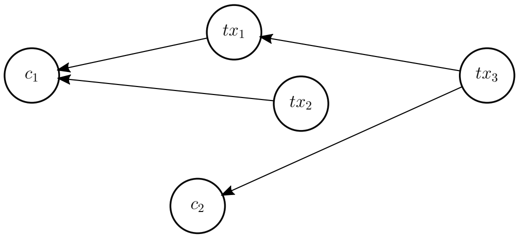

Figure 2. An example transaction graph with two coinbase transactions, c_1 and c_2, and three other, non-coinbase, transactions, tx_1, tx_2 and tx_3. The directed edges show the relations between transactions. In this example, tx_3 uses yet unspent outputs from tx_1 and c_2.

The graph cannot contain any directed cycles, because of how new transactions are appended to it. Whenever a new block is created, its transactions are added to the graph. The coinbase transaction does not depend on any transactions, hence only more recent transactions can use its outputs. Other transactions use the outputs of previous transactions, thus with every more recent transaction, at least one directed edge to an older transaction is introduced. Even if one transaction makes use of many previously unspent transaction at the time it is introduced, it only adds edges departing from it. It can therefore at that moment not be part of a directed cycle. As this is true for every new transaction, and since older transactions cannot be part of a cycle on their own, the newer transaction, to form a cycle, is required to be part of the cycle as well. As seen previously, this is not the case when the latter is added to the graph. The graph can therefore not contain any cycles. This property is important, because the computation of the mint centrality requires us to follow the directed edges to previous vertices (transactions) and eventually to coinbase transactions.

As discussed in Section 3, addresses are linked to transactions through the unspent transaction outputs. Every unspent transaction output belongs to an address that can spend it. Whether this address exists, or is in use is irrelevant. Thus, at every step, a property that is assigned to a transaction and its outputs can also be associated with the account that can use the output, if it has not yet been spent.

As a result, the mint centrality of an account is obtained as follows. We define the set of source coinbase heights S_{t,o_k} to contain the heights of all coinbase transactions that can be linked to an output o_k by following all outgoing edges from the transaction t that created the output o_k. The value k is used to identify the output within the transaction t.

- For the outputs of any coinbase transaction c, the height h_c of the block that contains the coinbase transaction is added. For every output o_k created directly by a coinbase transaction:

$$S_{c,o_k} = \{h_c\}$$

- For the output of any non-coinbase transaction t, we follow the outgoing edges to previous transactions t_p with outputs o_l, and obtain S_{t,o_k} through the union of all S_{t_p,o_l} of previous transactions, transactions whose output is used. This can be simplified, since for every transactions, it is indiscernible which output used which input.

$$S_{t,o_k} = \bigcup_{(t,t_p)\in E} S_{t_p,o_l} = \bigcup_{(t,t_p)\in E} S_{t_p}$$

Accordingly, the mint centrality of an address A at a given height h, is the number of distinct heights of coinbase transactions that it can be associated with through the transaction outputs it owned at any height prior to and including h. Since these are propagated through the graph G, the mint centrality mc can be expressed as follows.

$$mc(A,h) = \frac{|\bigcup_U S_{u_i}|}{h}$$

Where $U = \{u_1, u_2, \ldots, u_n\}$ is the set of all transactions u_i that created a transaction output owned by A, and n is the total number of these transactions. This shows how the mint centrality of an address is dependent on the height at which it is computed.

The mint centrality can be computed using a sparse matrix M with dimensions $N \times h$, where N is the number of known bitcoin address at height h. The elements of the Matrix M are the root-heights. We define the root-height $rh_j^{ad_i}$ to be either one or zero, representing whether or not the coinbase at height j is linked to address ad_i or not. The value i is simply used to distinguish between the different addresses ad_i. With this, the mint centrality $mc(ad_i, h)$ can be computed for any address as follows.

$$\frac{1}{h} \cdot M \cdot e = MC \iff$$

$$\frac{1}{h} \begin{bmatrix} rh_0^{ad_0} & rh_1^{ad_0} & \dots \\ \vdots & \ddots & \\ rh_0^{ad_N} & \dots & rh_h^{ad_N} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} mc(ad_0, h) \\ mc(ad_1, h) \\ \vdots \\ mc(ad_N, h) \end{bmatrix}$$

Where e is a vector of dimensions $h \times 1$ containing only ones and MC is the vector containing the values of the mint centralities for the currently known addresses. Consequently, the mint centrality can only assume values in the range $[0, 1]$.

Although we have until now considered the computation of the mint centrality between the heights 0 and h only, it is also possible to compute it between two chosen heights with a slight and trivial modification. Furthermore, whenever a new block is created, the matrix and vectors can easily be updated to represent the new state.

Finally, we consider the complexity of our implementation with which we compute the mint centrality.

First, the sparsity of the Matrix M is used to render the implementation more efficient. Second, the mint centrality can be computed along with the creation of new blocks. Whenever a block is created, the transactions it contains can be used to update M and thus the mint centrality. Also, the previous values can be reused. Hence, in order to obtain the mint centrality for every address, it is only necessary to go through the blockchain once. However, this requires us to keep track of unspent transaction outputs since these are the ones required to link transactions to previous transactions. Given the growth rate of unspent transaction outputs in the bitcoin network, this can become expensive. However, given that validating nodes already keep track of these unspent transaction outputs, an extension to accommodate our algorithm is simple.

5. Experimental Results

In order to evaluate the proposed centrality measure we computed the mint centrality for the first 200,000 blocks of the public bitcoin blockchain. We then inspected the 50 addresses with the highest mint centralities, and attempt to label them using online services such as *blockchain.info* and *bitcoinwhoswho.com.* This preliminary evaluation shows:

- In 18 out of the 50 addresses we found that they start with the hex value '1dice'. This belongs to the gambling service known as satoshidice. These are the addresses that receive the bets.
- In 25 out of the 50 addresses we found that they have transactions interacting with gambling services including satoshidice and luckybit.
- In the remaining 7 accounts we found additional well known addresses displaying a high mint centrality. Examples of these include: Gavin's original Bitcoin faucet[1] and the Wikileaks donation address[2].

[1] Gavin's original bitcoin faucet:'15ArtCgi3wmpQAAfYx4riaFmo4prJA4VsK'.

[2] Wikileaks donation address: '1HB5XMLmzFVj8ALj6mfBsbifRoD4miY36v'.

The results of our mint centrality seem to favour addresses related to the SatoshiDice gambling service [10]. Initially most bitcoin users and wallets would reuse the same address by default when creating a transaction, however it is now considered to be best practice to generate a new address to receive any return from a payment [11]. This is likely the reason why these addresses seem to bubble to the top of our mint centrality metric.

One of the features of bitcoin is the fact that it has a controlled bitcoin supply, with an eventual total amount of nearly 21 Million bitcoins expected to be reached by 2140 [1]. Although this implies that after that date no new bitcoins will enter the system, the core source of information for our algorithm, the coinbase transactions, will still be present. It will however be exclusively composed of the fees of the transactions included in that block.

6. Conclusions

In this work we have presented the 'mint centrality', a new centrality measure designed around the minting of new Bitcoins. We have shown an empirical analysis of a section of the bitcoin blockchain and presented the most central addresses. We identify most of them thanks to online explorers and thus show that the addresses obtained through our metric are well known.

In future work we will be performing an exhaustive comparison of our proposed centrality to other popular centrality measures utilising a larger section of the bitcoin blockchain.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] I. Ladegaard, "We know where you are, what you are doing and we will catch you: Testing deterrence theory in digital drug markets," *The British Journal of Criminology*, vol. 58, no. 2, pp. 414–433, 2017.

[3] M. Moser, R. Bohme, and D. Breuker, "An inquiry into money laundering tools in the bitcoin ecosystem," in *eCrime Researchers Summit (eCRS), 2013*. IEEE, 2013, pp. 1–14.

[4] M. Moser, "Anonymity of bitcoin transactions," 2013.

[5] A. Miller, J. Litton, A. Pachulski, N. Gupta, D. Levin, N. Spring, and B. Bhattacharjee, "Discovering bitcoin's public topology and influential nodes," 2015.

[6] S. Delgado-Segura, C. Pérez-Sola, G. Navarro-Arribas, and J. Herrera-Joancomartı, "Analysis of the bitcoin utxo set," in *5th Workshop on Bitcoin and Blockchain Research, Financial Cryptography and Data Security 18 (FC). Springer*, 2018. [Online]. Available: http://fc18.ifca.ai/bitcoin/papers/bitcoin18-final6.pdf

[7] P.-J. Kim and H. Jeong, "Reliability of rank order in sampled networks," *The European Physical Journal B*, vol. 55, no. 1, pp. 109–114, 2007.

[8] D. Koschützki and F. Schreiber, "Comparison of centralities for biological networks." in *German Conference on Bioinformatics*. Citeseer, 2004, pp. 199–206.

[9] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang, "Correlation between centrality metrics and their application to the opinion model," *The European Physical Journal B*, vol. 88, no. 3, p. 65, 2015.

[10] satoshidice, https://satoshidice.com//.

[11] D. Ron and A. Shamir, "Quantitative analysis of the full bitcoin transaction graph," in *Financial Cryptography and Data Security*, A.-R. Sadeghi, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 6–24.